DOCUMENT RESUME

ED 056 745 PS 005 029

AUTHOR Meyer, William J.; Egeland, Byron
TITLE Changes in Stanford-Binet IQ: Performance Vs Competence.
PUB DATE [68]
NOTE 13p.

EDRS PRICE MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS Caucasians; *Cognitive Ability; *Cognitive Development; Disadvantaged Groups; Inner City; Intelligence Differences; *Intelligence Quotient; *Intervention; Negroes; *Preschool Children; Summer Programs
IDENTIFIERS *Project Head Start; Stanford Binet

ABSTRACT

This evaluation of cognitive change in Head Start children focused on changes in performance as opposed to changes in competence; specifically, that Binet test performance improves as a function of experience with Binet examiners. The study involved 93 children assigned to four groups who were tested for IQ gains during a 6-week Head Start program in summer, 1968. Group distribution by sex was approximately equal; distribution by race was also equal in Groups I, II, and III. Group IV had only black children with pretest IQ that was meaningfully lower than that of the other groups. Groups were given the initial Stanford-Binet either at home or during the first or second week of school. At the end of the program, each group was again tested. A preliminary analysis of variance of the mean pretest IQ scores and the mean posttest scores showed that the four groups were essentially similar at the beginning and end of the program. These results indicate that, despite the possible variations in teachers and programs and in pretest IQ, overall differences among the groups were small and random indicating that test performance did not improve. Magnitude of change is also discussed in the study. [Not available in hard copy due to marginal legibility of original document]. (MK)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED056745

CHANGES IN STANFORD-BINET IQ. PERFORMANCE VS COMPETENCE[1]

William J. Meyer and Byron Egeland

The evaluation of changes in children's behavior as a result of experiences in Head Start classes, or other compensatory education programs, has typically involved a pre-test post-test design. Thus a single measure such as the Stanford-binet test, or a battery of generally correlated cognitive measures, are administered prior to the treatment experience and depending on the duration of the program (six week programs vs full year programs) the same battery is readministered and differences in performance are assessed. Appropriately, the t-test is used to evaluate the statistical significance of the change score. Since this statistical test is particularily sensitive to change, the results of the analyses are commonly statistically significant.

An important question of both theoretical and practical concern is the meaning that may be described to such findings. Since the original formulation of this study, Zigler and Butterfield (1968) published a paper focusing on the same issues of concern here (indeed the research design is very similar to the one used in this study). According to their theoretical analysis, with which we are in fundamental agreement, IQ changes may reflect: (1) Changes in formal cognitive processes, that is, changes in the level of cognitive functioning which we are labeling "competence". (2) Changes in information levels, that is, changes that reflect increased knowledge but not changes in level of cognitive functioning; and (3) "Motivational factors" which, from our viewpoint, reflect changes in the willingness to attempt items as opposed to passive nonresponding.

Aside from the Zigler and Butterfield study which provides support for

PS 005029

the motivational interpretation, there are no studies which have come to our attention focusing on the performance vs competence interpretation. Clearly Zigler et al. take the position that changes in motivation generate changes in performance level but not necessarily in cognitive levels. It is our hypothesis that the commonly reported IQ gains of between four and eight IQ points reflect motivational changes. It is our further hypothesis that these motivational changes are reflected directly in performance rather than in competence changes.

The general design of this study involves four groups of children who received a pre- and a post-test at different points during a six week summer Head Start Program; this phase of the research design relates to the motivational interpretation. The second component of this project involves an item analysis of the Stanford-Binets for each of the 100 children in the sample. These analyses will focus on the nature of the pre-test items passed and the post-test items passed and will examine the nature of the post-test items passed. In performing this latter analysis it is our assumption that certain of the Binet items appearing at the six-year level are indicative of higher cognitive levels and thus may reflect changes in level. This first report is related to the impact of the initial testing experience on changes in motivation.

METHOD

Subjects The children in this study were enrolled in a six week Head Start Program during the Summer, 1968. All the children resided in what is generally known as the "inner city". Of the original sample of 100 children, two children were unavailable for pre-testing and an additional five children were not available, largely because they moved away, for the

post-testing. Table 1 shows the Ns and CAs by sex race, and treatment group.

Unfortunately, it was not possible to randomly assign children to each of the treatment groups. A preliminary analysis of the data indicated, however, that the differences between the groups on pre-test IQ are not statistically significant ($F = 2.4$; $df = 3 \& 94$; $p > .05$). The distribution by sex for all four groups is approximately equal as is the distribution for race in Groups I, II, and III. Group IV is comprised entirely of Negro children and despite the failure to attain statistical significance it would appear that the Group IV pre-test IQ is perhaps meaningfully lower than that of the other groups. The significance of these sampling differences will be discussed later in this paper in context with the data.

Procedure. There were a total of four groups which varied in terms of when they were given the initial Stanford-Binet during the six week Head Start Program. Group I was initially tested at home during the week prior to the beginning of the Head Start Program. It was thought that perhaps administering the test in familiar surroundings might minimize the impact of having to work with a strange and somewhat demanding adult. This group was post-tested during the last week of the Head Start Program. Group II was administered the Binet during the first week of the Head Start Program and during the last, or sixth week. This group is considered the standard or control group. Group III was administered the Stanford-Binet during the second week and again during the sixth week. It was hypothesized that this group might benefit from one week of encounters with strangers and would thus show higher pre-test performance and conceivably a lower change score. Group IV was administered the Stanford-Binet during the first week,

TABLE I

Sample Characteristics

Ns and CAs by Race and Sex

Groups

	I				II				III				IV			
	Male		Female		Male		Female		Male		Female		Male		Female	
	N	W	N	W	N	W	N	W	N	W	N	W	N	W	N	W
N	4	7	9	5	4	12	3	3	6	6	9	4	10	0	14	0
MCA(Mos)	62.75	60.86	62.44	61.20	62.75	64.33	65.00	59.66	68.17	62.00	65.11	66.75	62.0		64.4	
SDca	4.16	2.95	3.30	3.50	3.70	3.50	3.21	3.86	3.70	3.21	5.13	4.71	5.0		5.4	

the second week, and the sixth week. This is the critical group in our design because it can show the effects of experience with examiners (week one vs week two) and can further indicate if there are meaningful practice effects from repeated testing (week two vs week six).

All children were administered the full-length Stanford-Binet by examiners experienced in working with Head Start children. Standard testing procedures were employed; that is, no specific efforts were made to maximize the children's performance.

The Nursery School Programs in which the children were enrolled can most appropriately be described as traditional in approach. This implies that the children were provided with standard nursery school equipment, such as blocks, dress-up corners, and other toys, and a program involving such things as reading stories, coloring and pasting, and field trips. Each classroom was comprised of a head teacher trained in early childhood education and a teacher-aide who were mainly from the neighborhood. (These comments on the nature of the program are derived entirely from general impressions since we made no effort to document program content in detail.)

Results

A preliminary analysis of variance of the mean pre-test IQ scores and the mean post-test IQ scores indicates that the four groups are essentially similar at the beginning and at the end of the programs ($F = 2.4$; $df = 3$ & 94, $p > .05$; $F = 2.0$ $df = 3$ & 90 $p > .05$, pre- and post-tests, respectively). These results indicate that despite the possible variations in teachers and programs and despite the variation in pre-test IQ, the overall differences among the groups at the conclusion of the six-week program are small and random.

Next we examined the magnitude of change within each of the four treatment groups. Table 2 summarizes the means and SDs of the pre- and post-test means and the mean change scores for each of the groups. A series of matched t tests to determine the significance of the change scores were run for each group: Group I, $t = 4.2$ $df = 22$. $p = < .01$; Group II, $t = 2.8$ $df = 24$, $p < .01$; Group III, $t = 3.9$; $df = 24$. $p = < .01$. Group IV analysis involved a within groups analysis of variance comparing Week 1 vs Week 2 vs Week 6. $F = 6.4$. $df = 2\&2$: $f < .01$). Multiple comparisons were run between Week 1 and Week 2 and Week 2 and Week 6 ($t = 3.6$; $df = 22$. $p < .01$ and $t = .2$. $df = 22$, $p = > .05$ respectively). These results indicate that for Groups I, II, and III, the average gain in IQ score is statistically significant and is of a magnitude commensurate with that frequently reported in other studies (see the control group gains in the Zigler and Butterfield study, for example). It would appear, therefore, that nothing unusual is occurring in our sample of children or classroom teachers. Of particular pertinence to this study, however, is the fact that in Group IV the change in IQ between Week 1 and Week 2 is statistically significant and of the same magnitude as the other groups. The change between Week 2 and Week 6 is not only not statistically significant but the children actually showed a very slight drop in performance. These results are consistent with our hypothesis, and that proposed by Zigler and Butterfield (1968), that the typical gains reported in IQ may well be a function of experience with examiners as contrasted to the experiences gained in Head Start classrooms.

The next phase of the analyses compared the difference scores for each of the four groups. In the first between-groups analysis, the difference

TABLE II

Means and Standard Deviations

Pre-Test, Post-Test, Change Scores

Groups

	I		II		III		IV	
Week	M	SD	M	SD	M	SD	M	SD
Home	84.4	13.5	-	-	-	-	-	-
1	-	-	83.4	14.0	-	-	77.0	15.9
2	-	-	-	-	88.7	17.4	81.7	15.5
6	90.6	12.9	88.0	14.1	92.1	18.4	81.4	16.7
Change (6-2)	6.2	7.3	4.6	7.9	3.4	5.5	4.7 (-.3)	6.2 (6.4)

score included for Group IV was the difference between Week 1 and Week 2 and the second between-groups analyses of variance involved the difference between Week 2 and Week 6. The results of these analyses were: $F = .44$, $df = 3 \& 89$, $p = > .05$ and $F = 3.99$, $df = 3 \& 89$, $p = < .05$, respectively. The results of the first analyses of variance suggest that the effects of testing the children at home, delaying the first administration of the Binet until the second week, and the effect of post-testing after only one week, had no significant effects on change scores. These results can tentatively be interpreted as meaning that familiarity of surroundings and one week of experience with adults in a school setting does not particularly influence pre-test performance. This data further indicate that extended (5 week) experiences in a program does not effect changes. This conclusion is further supported by the results of the analyses of variance of the pre-test scores, previously reported, in which there were no statistically significant differences between the four groups.

Of particular interest is the result of the second analyses of variance in which the change score for Group IV involved the difference between performance on the test during Week 2 and during Week 6. In considering these results it should be noted that the mean difference for Group II and the mean difference for Group IV, taking the difference in Group IV as Week 1 vs Week 6, are essentially identical. A comparison of the change scores for Group III (tested Week 2 and Week 6) with the comparable period of time for Group IV reveals a statistically significant difference ($t = 2.67$, $df = 44$, $p = < .01$). Differences between change scores for Group I vs Group II and Group II vs Group III are not statistically significant ($t = .85$, $df = 45$; $p = > .05$; and $t = .03$; $df = 45$ $p = > .05$, respectively).

The results of these analyses suggest rather strongly that the observed magnitude of change in IQ performance in Groups I through III occurs as a result of the initial examination and not as a result of time in the Head Start situation.

In addition to the analyses concerned with the major issues in this study, additional analyses were performed which may be of some interest. It was possible, for example, to examine sex and race differences on both the pre-test and post-test scores as well as the magnitude of change scores. A two by two analysis of variance of the pre-test scores indicated that the sex difference is not statistically significant ($F = 3.4$ $df = 1$ & 96, $p > .05$) that the race difference is not statistically significant ($F = 3.0$, $df = 1$ & 96; $p > .05$); and sex x race interaction is not statistically significant ($F = 1.2$, df 1 & 96; $p = > .05$). Performance on the post-test indicated only a statistically significant sex difference ($F = 5.1$; df 1 & 91, $p = < .01$). Inspection of the pre-test data indicates that the direction of the sex difference is in favor of the males; whereas the direction of the race difference is in favor of the white children. The same directional differences occur for the post-test but these differences achieve statistical significance only in the case of sex. In view of the fact that the recruiting practices employed in making up the various groups are unknown, it is not particularly clear why the male subjects, and this was true for both racial groups, performed better on the Stanford-Binet than the females.

A frequently reported finding is a correlation between the pre-test score and the magnitude of the change score. Typically, this correlation is in a negative direction indicating that children with lower initial scores tend to gain somewhat more on the post-test. In the context of this study,

the negative correlation could be interpreted as indicating that children with lower pre-test scores are more adversely affected by their initial interaction with the examiner and, conversely, benefit more from the initial examination experience than children with higher pre-test scores. In this study none of the correlations between pre-test performance and change score were statistically significant. This result suggests that regardless of initial test performance, the children benefited in a uniform fashion from the original experience with the examiner.

Discussion

The purpose of this phase of the project concerned with the evaluation of cognitive change in Head Start children focused on changes in performance as opposed to changes in competence. More specifically, this study examined the hypothesis that Binet test performance improves as a function of experience with Binet examiners. As part of the overall design, this study examined the effects of pre-testing at home in a presumably familiar environment where the children would perhaps be more responsive and also examined the effects of delaying the initial examination for a week, giving the children the experience with strange adults. The major outcome of the study was that as much gain in performance occurred in one week as occurred over a period of at least five weeks. The effect of testing at home and of delaying the initial examination for one week were not only not statistically significant but did not influence average performance to any degree whatsoever.

Precise and unequivocal interpretations of the data derived from this study are impossible for at least three reasons: (a) the children were not randomly placed in the different treatment groups; (b) there was no control

group that had been pre-tested and post-tested without an intervening experience. (c) the duration of the program was only for a period of six weeks. In the case of the first limitation, at least three supportive arguments can be made: (1) despite the lower pre-test IQ of Group IV, all four groups evidenced essentially the same amount of gain and this gain (4.6 IQ points) is essentially similar to that reported by Zigler and Butterfield (1968.) (2) the correlation between pre-test score and change score is not statistically significant for any of the groups--suggesting that pre-test variation did not influence magnitude of change. (3) it is unclear what subject characteristics might exist that would allow for a 4.6 gain in IQ between the first and second testing, but a loss of .3 IQ points between the second and third testing. It is conceivable, of course, that some variables are operating and this might well be worth examining in a truly randomized design.

With respect to the third limitation of this study, namely the six-week duration, it should be noted that the mean change found in this study is commensurate with the outcome of full-year Head Start programs. From the basic viewpoint adopted in this study, this outcome is not surprising. As noted earlier, the frequently reported significant negative correlation between pre-test performance and change score, suggests that in those studies reporting larger mean gains it is conceivable that the subjects had lower initial test scores. It is a further contention, however, that the magnitude of IQ change may be a relatively unimportant variable. Thus, even an IQ change of as much as 10 or 12 points may not, depending upon the initial test score, signify any meaningful change in level of cognitive functioning. The degree to which IQ gain does reflect shifts in cognitive level is currently

being examined in our laboratory.

A comment or two seems in order on the sample of children initially tested at home. Recall that it was our hypothesis that these children would perhaps show a smaller gain score because their pre-test scores would not have been adversely influenced by strange surroundings. Clearly, this did not occur. Indeed, it almost appears as if these children were adversely affected by having their initial examination at home. The junior author did the initial home examination and according to his reports there were substantial distractions at home in terms of general noise level deriving both from adults and from other children. Thus, it would appear that the procedure of examining children at home is not feasible.

Finally, these data cannot be interpreted as indicating anything particularly negative about the four Head Start classes that were employed. Our observations indicated that these classrooms were not atypical of other Head Start classes in our experience and that the teachers were both well-trained and committed to the program. It should also be noted that we have little information to determine why some children showed larger gains than others, yet this variation in change score certainly suggests that there were attributes in each of these classrooms that positively influenced at least some children. If anything, the results of this study in conjunction with those reported by Zigler and Butterfield suggest that a considerable amount of basic work is needed before we will gain an understanding of cognitive development and thus be in a position to develop better programs. One implication of these data, for other investigators, is to consider the magnitude of IQ change scores in their own studies in terms of the amount of IQ gain to be anticipated on the basis of experience with examiners.

Perhaps the most important implication, however, is the need for a shift in strategy from reporting change scores to a strategy in which we intensively investigate the characteristics of children as they interact with the characteristics of the programs--an often stated objective that somehow has not yet been implemented.